09/873,340

DIALOG(R)File 351:Derwent WPI

011999278 **Image available**
WPI Acc No: 1998-416188/199836
Related WPI Acc No: 1998-416315; 1998-416316
XRPX Acc No: N98-324033

Fibre channel attached storage architecture for computer network - transfers data between server and storage device at minimum data rate of 100 megabytes per second, storage device is hot pluggable into Fibre Channel while computer system is powered

Patent Assignee: COMPAQ COMPUTER CORP (COPQ)
Inventor: ALEXANDER D J; CALLISON R A; FREEMAN E E; GALLOWAY W C; GRANT D L ; GRIEFF T W; MCCARTY J F; RUSHTON T D; SABOTTA M L; SCHNEIDER R D; SCHOLHAMER G J; SCHULTZ S M; SKIDMORE A E; THOMPSON M J; GUNLOCK R D; MCGOWEN M E

Number of Countries: 026 Number of Patents: 006
Patent Family:

Patent No	Kind	Date	Applicat No	Kind	Date	Week	
EP 858036	A2	19980812	EP 98300808	A	19980204	199836	B
JP 10240670	A	19980911	JP 9829897	A	19980212	199847	
JP 10243020	A	19980911	JP 9828328	A	19980210	199847	
JP 10293633	A	19981104	JP 9843374	A	19980225	199903	
US 5954796	A	19990921	US 97798962	A	19970211	199945	
US 6014383	A	20000111	US 97797129	A	19970210	200010	

Priority Applications (No Type Date): US 97805281 A 19970225; US 97797129 A 19970210; US 97798962 A 19970211
Cited Patents: -SR.Pub
Patent Details:

Patent No	Kind	Lan	Pg	Main IPC	Filing Notes
EP 858036	A2	E	13	G06F-013/42	
Designated States (Regional): AL AT BE CH DE DK ES FI FR GB GR IE IT LI LT LU LV MC MK NL PT RO SE SI					
JP 10240670	A		11	G06F-013/14	
JP 10243020	A		10	H04L-012/56	
JP 10293633	A		10	G06F-003/00	
US 6014383	A			H04B-010/08	
US 5954796	A			G06F-009/00	

Abstract (Basic): EP 858036 A

The system includes a Fibre Channel interconnection and a server and a storage device are connected to the Fibre Channel a host bus adapter connected between the server and the Fibre Channel. A media module connected between the server and the Fibre Channel. An array controller connected between the Fibre Channel and the storage device. A redundant Fibre Channel connected to the server and connected to the storage device. The storage device is hot pluggable into the Fibre Channel while the computer system is powered. Data can be transferred between the at least one server and the at least one storage device at a minimum data rate of 100 megabytes per second.

ADVANTAGE - Eliminates use of SCSI bus and cabling which can be heavy messy and inflexible.

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平10－243020

(43)公開日　平成10年(1998)９月11日

(51)Int.Cl.[6]	識別記号	ＦＩ	
Ｈ０４Ｌ　12/56		Ｈ０４Ｌ　11/20	１０２Ｚ
Ｇ０６Ｆ　13/00	３５７	Ｇ０６Ｆ　13/00	３５７Ｃ
Ｈ０４Ｌ　12/28		Ｈ０４Ｌ　11/00	３１０Ｄ

審査請求　未請求　請求項の数８　ＯＬ　（全 10 頁）

(21)出願番号　特願平10－28328

(22)出願日　平成10年(1998)２月10日

(31)優先権主張番号　７９７１２９

(32)優先日　1997年２月10日

(33)優先権主張国　米国（ＵＳ）

(71)出願人　591030868
コンパック・コンピューター・コーポレーション
COMPAQ COMPUTER CORPORATION
アメリカ合衆国テキサス州77070，ヒューストン，ステイト・ハイウェイ　249，20555

(72)発明者　ジェームズ・エフ・マカーティ
アメリカ合衆国テキサス州77379，スプリング，ワンズワース・ドライブ　9227

(74)代理人　弁理士　社本　一夫　（外５名）

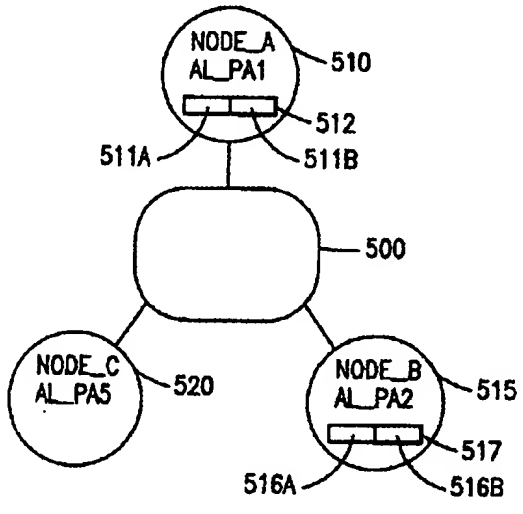
最終頁に続く

(54)【発明の名称】　ファイバ・チャネル環境において多数のイニシエータを制御するシステムおよび方法

(57)【要約】

【課題】　コンピュータ・システムにおいてファイバ・チャネル（ＦＣ）通信環境を管理して、伝送中段が生じないようにする。

【解決手段】　コンピュータ・システムはＦＣプロトコルと互換性のある複数のＦＣデバイス５１０，５１５，５２０を備え、該デバイスの内少なくとも２つはイニシエータ５１０，５１５であり、他のデバイスはターゲット（宛先ノード）５２０である。イニシエータ及びターゲットは調停されたループ５００のＦＣ環境を構成し、またイニシエータは、記憶アレイである未決リンク・サービス・アレイ５１２、５１７を備えている。イニシエータは、該アレイに他のイニシエータ又はターゲットに送るための要求フレームに関連するタイプ情報を記憶し、その後、該要求フレームを送信する。受信側では要求フレームに応答して、応答フレームのタイプ情報を記憶することなく、応答フレームを送出する。



【特許請求の範囲】
【請求項1】　複数のファイバ・チャネル（FC）・デバイスを含み、該FCデバイスの内少なくとも2つはイニシエータであるファイバ・チャネル通信環境において、該ファイバ・チャネル通信環境を管理しかつ制御する方法であって、
要求フレームに関連するタイプ情報要素を記憶した後に、前記少なくとも2つのイニシエータの各々から、前記ファイバ・チャネル・デバイスの各々に前記要求フレームを送信するステップと、
受信した要求フレームに応答して、応答フレームのタイプ情報要素を記憶することなく、前記少なくとも2つのイニシエータの各々から、前記応答フレームを送出するステップとから成ることを特徴とする方法。
【請求項2】　請求項1記載の方法において、該方法は更に、仲裁ループに関連するループ初期化プロセスを完了するステップを含むことを特徴とする方法。
【請求項3】　請求項1記載の方法において、該方法は更に、前記要求フレームに応答して受信した応答フレームに関連するペイロードの内容を操作することなく、前記少なくとも2つのイニシエータの各々によって、前記応答フレームを処理するステップを含むことを特徴とする方法。
【請求項4】　複数のファイバ・チャネル（FC）・デバイスを含み、該ファイバ・チャネル・デバイスの内少なくとも2つがイニシエータであるファイバ・チャネル通信環境において、該ファイバ・チャネル通信環境を管理しかつ制御するシステムであって、
要求フレームに関連するタイプ情報要素を記憶した後に、前記少なくとも2つのイニシエータの各々から、前記ファイバ・チャネル・デバイスの各々に前記要求フレームを送信する手段と、
受信した要求フレームに応答して、応答フレームのタイプ情報要素を記憶することなく、前記少なくとも2つのイニシエータの各々から、前記応答フレームを送出する手段とから成ることを特徴とするシステム。
【請求項5】　請求項4記載のシステムにおいて、該システムは更に、仲裁ループを初期化するループ初期化手段を備えていることを特徴とするシステム。
【請求項6】　請求項4記載のシステムにおいて、該システムは更に、前記要求フレームに応答して受信した応答フレームに関連するペイロードの内容を操作することなく、前記受信した応答フレームを処理する手段を備えていることを特徴とするシステム。
【請求項7】　少なくとも2つのイニシエータを含む仲裁ループ・トポロジを処理する方法であって、
ループ初期化ステップを実行するステップと、
前記少なくとも2つのイニシエータの各々によって送出される要求フレームに関連するタイプ情報要素を記憶するステップと、
前記少なくとも2つのイニシエータの各々から要求フレームを送出するステップと、
前記少なくとも2つのイニシエータの各々から応答フレームを送出するステップであって、前記要求フレームに応答して前記応答フレームを送出するステップとから成ることを特徴とする方法。
【請求項8】　請求項7記載の方法において、該方法は更に、前記要求フレームに応答して受信した応答フレームに関連するペイロードの内容を操作せずに、前記少なくとも2つのイニシエータの各々によって、前記応答フレームを処理するステップを含むことを特徴とする方法。
【発明の詳細な説明】
【0001】
【発明の属する技術分野】本発明は、チャネルおよびネットワーク通信システム並びにチャネルおよびネットワーク通信方法に関し、更に特定すれば、ファイバ・チャネル環境において多数のイニシエータ(initiator)を制御するシステムおよび方法に関するものである。
【0002】
【従来の技術】デバイス通信には、チャネルおよびネットワークという2種類のプロトコルがある。チャネルは、マスタ・ホスト・コンピュータおよびスレーブ周辺デバイス間において、一旦データ通信が開始したなら、ソフトウエアのオーバーヘッドがほとんどなく比較的短い距離において超高速で大量のデータを搬送するように設計されている。チャネルは、通常、ハードウエア集約的であり、マスタおよびスレーブ間に直接接続即ち切替式二点間接続を提供する。一方、ネットワークは、中距離ないし長距離において複数のホストおよびシステム資源を共有する多くのユーザにインターフェースし、多くのトランザクションに対応するように設計されている。ネットワークに関しては、高い接続性が得られる限り、通常オーバーヘッドの増大は容認することができるものである。
【0003】ファイバ・チャネル・プロトコル（「FCP：Fiber Channel Protocol」）は、これら2つの異種通信方法の最良点を組み合わせて、単一のオープン・システム・インターフェース状（OSI状）のスタック・アーキテクチャとする、新世代のプロトコルである。本質的に、ファイバ・チャネル（「FC：Fiber Channel」）は、マルチ・トポロジ(multi-topology)の多層スタックであり、物理的搬送特性を制御する下位レイヤ・プロトコル（「LLP：lower-layer-protocol」）、および/またはオペレーティング・システムと互換性のある上位ソフトウエア構造との間でLLP通信のマッピングを行う上位レイヤ・プロトコル（「ULP：upper-layer-protocol)を備えている。これらのULPは、チャネル・プロトコルおよびネットワーク・プロトコルの双方を含み、例えば、インテリジェント周辺インターフェース

（ＩＰＩ：Intelligent Peripheral Interface）、スモール・コンピュータ・システム・インターフェース（「ＳＣＳＩ：Small Computer System Interface）、およびインターネット・プロトコル（「ＩＰ：Internet Protocol」）のようなプロトコルを含んでいる。

【０００４】チャネル通信またはネットワーク通信のいずれかに関与するデバイスは、「イニシエータ」または「ターゲット」、あるいはこれらの双方として分類されることは公知である。いくつかの特定の機能がイニシエータまたはターゲットのいずれかに割り当てられる。すなわち、（ｉ）イニシエータは、通信経路の仲裁（アービトレーション）を行い、ターゲットを選択することができる。（ｉｉ）ターゲットは、コマンド、データ、ステータス、またはその他の情報のイニシエータへの転送、若しくはイニシエータからの転送を要求することができる。（ｉｉｉ）場合によっては、ターゲットが通信経路の仲裁（調停）を行い、トランザクションを継続するためにイニシエータを選択し直すことも可能である。

【０００５】ファイバ・チャネル・プロトコルと共に動作可能なデバイスでは、イニシエータ機能を有するデバイスのみが、当技術ではリンク・サービス要求(Link Service Request)または拡張リンク・サービス要求(Extended Link Service Request)として知られている要求を開始することができる。リンク・サービス・コマンドは、ファイバ・チャネル・イニシエータに、ノード発見、要求中止、および通信フレーム拒絶というようなタスクを実行する能力を与える。ファイバ・チャネル・ターゲットが開始できる唯一のリンク・サービス・コマンドは、コマンド／フレームの拒絶（「ＬＳ＿ＲＪＴ」）である。

【０００６】一般に、単一のイニシエータＦＣ環境では、イニシエータ・デバイスは、必要とされるリンク・サービス・コマンドを送出し、それに応答してターゲットから送られてくるリンク・サービス承認（「ＬＳ＿ＡＣＫ」）フレームまたはリンク・サービス拒絶フレーム（ＬＳ＿ＲＪＴ）の受信を待機する。以降、これらのＬＳ＿ＡＣＫおよびＬＳ＿ＲＪＴフレームのことを、総称して、「応答」フレームと呼ぶことにする。一方、マルチ・イニシエータ環境(multi-initiator environment)では、イニシエータは、リンク・サービス・コマンドの受信側および送出側双方として動作する。これら二重の役割のために、かかるイニシエータは、応答フレームの受信側および送出側の双方として動作する。

【０００７】

【発明が解決しようとする課題】イニシエータ間におけるこれら双方向伝送のために、マルチ・イニシエータＦＣ環境においては、激しい混乱が発生する可能性が非常に高い。実際、イニシエータ側の混乱によって、実行可能な通信リンクを確立する役割を担っている初期化手順が停止(stall)し、それにより、マルチ・イニシエータ環境において送信中断が発生してしまう。１つのチャネルまたはネットワーク通信システム内に多数のイニシエータを設けることは、概念的には処理能力レベルを高め、接続度(degree of connectivity)上昇を達成するであろうが、マルチ・イニシエータＦＣ環境において上述のような中断が発生すると、通信リンクのダウン・タイムが高性能の最先端システムでは容認できないレベルにまで上昇してしまう可能性がある。

【０００８】種々の単一イニシエータによるＦＣ実施態様は存在しているが、上述した問題や欠陥に適切に対処し、以下に説明する本発明の利点および新規の特徴全てを有するマルチ・イニシエータＦＣ通信システムは、未だ提供されていない。

【０００９】

【課題を解決するための手段】本発明は、コンピュータ・システムにおいてファイバ・チャネル（ＦＣ）通信環境を管理しかつ制御する方法を提供することによって、先に確認した問題、および既存の技術のその他の欠点や欠陥を克服するものである。かかる環境は、複数のＦＣデバイスを含み、少なくともその内２つのＦＣデバイスが、イニシエータである。この管理および制御方法は、要求フレームに関連するタイプ情報要素を記憶した後に、イニシエータの各々から、複数のＦＣデバイスの各々に要求フレームを送信するステップと、受信した要求フレームに応答して、応答フレームのタイプ情報要素を記憶せずに、少なくとも２つのイニシエータの各々から、応答フレームを送出するステップとを含んでいる。

【００１０】本発明は、更に、コンピュータ・システムにおいて、多数のイニシエータを含む複数のファイバ・チャネル（ＦＣ）デバイスを備えたＦＣ通信環境を管理しかつ制御するシステムを提供する。このシステムは、要求フレームに関連するタイプ情報要素を記憶した後に、イニシエータの各々から、複数のＦＣデバイスの各々に前記要求フレームを送信する手段と、受信した要求フレームに応答して、応答フレームのタイプ情報要素を記憶せずに、イニシエータの各々から、応答フレームを送出する手段とから構成されている。本発明のより完全な理解は、添付図面と関連付けて以下の詳細な説明を参照することによって得ることができよう。

【００１１】

【発明の実施の形態】これより図面を参照して説明するが、図面全体にわたって同様または類似する要素には同一の参照番号を付している。また、種々の要素は必ずしも同一の拡縮率で描かれている訳ではない。

【００１２】図１には、本発明を実施可能な、一例のコンピュータ・システム２００のブロック図が示されている。当業者には明らかであろうが、コンピュータ・システム２００は、ここではその機能的なブロックとして表されている。オペレーティング・システム（「ＯＳ」）２１０が、コンピュータ・システム２００内に設けら

れ、それに関連する情報の流れを制御する。ＯＳ２１０は、ディスク・オペレーティング・システム（「ＤＯＳ：Disk Operating System」）、あるいは、コンピュータ・システム２００をネットワーク環境内に配置するか否かに応じて、例えば、Windows NT（登録商標）またはNetWare（登録商標）のような、適切なネットワーク・オペレーティング・システム（「ＮＯＳ：Network Operating System」）とすることができる。

【００１３】更に、ＯＳ２１０は、例えばＳＣＳＩ規格のような、少なくとも１つの従来からのチャネル通信インターエースと共に動作可能である。例示のＯＳ２１０には、更に、例えば、Internet Protocol（「ＩＰ」）のような、従来のネットワーク通信プロトコルとの相互動作を可能にするような機能構造を設けることも可能である。

【００１４】引き続き図１を参照すると、例示のＯＳ２１０は、上位通信経路２３０を通じて、ＯＳ互換チャネルまたはネットワーク通信プロトコル／インターフェース２１５と通信を行う。例示のコンピュータ・システム２００の機能ブロック図において、上位通信経路２３０は、例えば、ＳＣＳＩプロトコル・ドライバまたはＩＰプロトコル・ドライバといった、通信プロトコル・ドライバのようなＯＳソフトウエア構造を含んでいる。例示のＯＳ２１０およびＯＳ互換インターフェース／プロトコル２１５は、コンピュータ・システム２００において一体となっており、、これ以降、ＯＳ環境２５０と呼ぶことにする。参照番号２２０は、ファイバ・チャネル（「ＦＣ」）環境を示し、以下で更に詳しく説明する公知のファイバ・チャネル・プロトコル（「ＦＣＰ」）アーキテクチャに加えて、本発明の教示にしたがって動作可能な複数のＦＣデバイスを含んでいる。

【００１５】更に図１を参照すると、例えば、ＯＳ２１０を含む殆どのオペレーティング・システムには、ＦＣ環境２２０内に配置されているデバイスと「直接」通信する機能が備えられていないことが明らかであろう。したがって、例示のコンピュータ・システム２００において、ＦＣ環境２２０の利点を奏するように構成し利用するために、ＦＣ環境２２０およびＯＳ互換通信インターフェース２１５の間に、リンク経路２２５を設けている。

【００１６】次に図２を参照すると、図２には、ＦＣＰスタック・アーキテクチャ３００の模式図が示されている。容易に認めることができようが、ＦＣＰアーキテクチャは、オープン・システム・インターフェース（「ＯＳＩ」）スタックと全く同様に、プロトコル・レイヤの階層的集合として構成されている。ＦＣスタックの下位３レイヤ（ＦＣ－０～ＦＣ－２で示すレイヤ３１０～レイヤ３２０）が、ファイバ・チャネル物理規格（「ＦＣ－ＰＨ：Fiber Channel Physical Standard」）として知られている物理規格を形成する。この規格は、例えば、ＦＣ環境２２０（図１に示した）を含む、ファイバ・チャネル環境の物理的伝送特性全てを規定するものである。残りのレイヤ（ＦＣ－３で示すレイヤ３２５、およびＦＣ－４で示すレイヤ３３０）は、他のネットワーク・プロトコルおよびアプリケーションとのインターフェースを処理する。イーサネットやトークン・リングのような既存のローカル・エリア・ネットワーク（「ＬＡＮ：Local Area Network」）とは異なり、ＦＣは、スタック３００の種々の機能レイヤを、物理的に分離して保持している。この物理的な分離によって、いくつかのスタック機能をハードウエアで実施し、その他の機能をソフトウエアまたはファームウエアで実施することが可能となることは明らかであろう。

【００１７】レイヤ３１０、即ちＦＣ－０は、ＦＣアーキテクチャの最下位の機能レイヤであり、ＦＣ環境２２０（図１に示す）に配置された複数のＦＣデバイス間のリンク接続の物理的特性を記述する。ＦＣ－０は、１３３Ｍｂａｕｄ（ボー）の基準速度、最も一般的に使用されている速度である２６６Ｍｂａｕｄ、ならびに５３１Ｍｂａｕｄおよび１．０６２Ｇｂａｕｄに対応する。しかしながら、リンク接続を確立し維持する際に伴うオーバーヘッドのために、実際のデータ・スループットはいくらか低く、１３３Ｍｂａｕｄでは１００Ｍｂｉｔ／ｓ、２６６Ｍｂａｕｄでは２００Ｍｂｉｔ／ｓ、５３１Ｍｂａｕｄでは４００Ｍｂｉｔ／ｓ、そして１．０６２Ｇｂａｕｄでは８００Ｍｂｉｔ／ｓとなる。更に、ＦＣ－０は、単一モードまたはマルチモードの光ファイバ・ケーブル、同軸ケーブル、およびシールド・ツイスト線対（「ＳＴＰ：shielded twisted pair」）媒体を含む、広範囲の物理的ケーブルに対応する。これらのケーブル要素の各々は、ある範囲のデータ速度に対応し、特定の距離制限を賦課するが、ＦＣは，図２に示したＦＣ環境２２０のような同一ＦＣ環境内において、これらの全てを混合することができる。例えば、単一モードｎｏ光ファイバを１０ｋｍまでの距離に使用することができ、２００Ｍｂｉｔ／ｓのマルチモード・ファイバを２ｋｍまでの距離に使用することができ、１００Ｍｂｉｔ／ｓに対応するＳＴＰを５０メートルまでの距離に使用するようにしてもよい。

【００１８】レイヤ３１５、すなわちＦＣ－１は、直列エンコード（符号化）およびデコード（複合化）規則、特殊特性、ならびにエラー制御を含む、伝送プロトコルを定義する。ＦＣ－１は、８Ｂ／１０Ｂブロック・コードを用い、８データ・ビット毎に、ディスパリティ・コントロール(disparity control)として知られている、エラー検出および訂正のための２つの余分なビットを付け加え、１０ビット群として送信する。８Ｂ／１０Ｂ方式は、十分なエラー検出および訂正機能を与えるので、低コストのトランシーバを用いることができ、タイミング復元方法を用いて、無線周波数干渉の危険性を低下さ

せると共に、均衡の取れた同期送信を確保することができる。

【0019】FC－PHの第3レイヤ320、即ちFC－2は、FCデバイス間でデータがどのように転送されるかについて記述し、各FCデバイスは「ノード」に配置され、フレーム・フォーマット、フレーム・シーケンス、通信プロトコル、およびサービス・クラスの定義を含む。ファイバ・チャネルにおけるデータ送信の基本単位は、可変サイズのフレームである。フレームの長さは、2，148バイトまでとすることができ、2，048バイトまでのペイロード、フレーミング(framing)、送信元（ソース）および宛先ポート・アドレシング、サービス・タイプ、ならびにエラー検出情報を与える36バイトのオーバーヘッド、ならびにユーザ・データ、即ち、ペイロードについてのその他の種々雑多な情報のための64バイトの付加的な任意のオーバーヘッドから成る。単一の上位レイヤ（即ち、スタック300における上位レイヤ）のプロトコル・メッセージは、フレームのペイロード容量よりも大きくすることができ、その場合、メッセージは、シーケンスと呼ばれる、一連の関連付けられたフレームに分割される。

【0020】引き続き図2を参照すると、FC－2レイヤは、FCPスタック300の主要な「ワークホース(workhorse)」であり、FC－0レイヤを通じた送信のために、上位レイヤ（レイヤ325，330）からのデータをフレーム化して、連続的に送り出す。また、FC－0レイヤからの送信を受け入れ、上位レイヤ325，330による使用のために、必要であれば、それらを再度フレーム化し、そして再度連続的に送り出す。2つのノード間の全二重送信経路を定義することに加え、FC－2レイヤは、フロー・コントロール（制御）、リンク管理、バッファ・メモリ管理、ならびにエラー検出および訂正を含む、必須のトラフィック管理機能も提供する。

【0021】FCPスタック300の重要な特徴の1つは、FC－2レイヤが4つのサービス・クラスを定義し、種々の通信の要望に答えることである。クラス1サービスは、専用の無中断通信リンクである、ハード・ワイヤ、即ち、回路切替型接続を定義する。このサービスは、その期間の間、接続の排他的使用（時として「利己的接続」と呼ばれる）を提供する。クラス1サービスは、2台のスーパーコンピュータ間のように、時間に厳しく「バーストを発生させない」専用リンクのために設計されたものである。クラス2サービスは、配信を保証し、トラフィックの受信を確認する、無接続フレーム交換送信である。フレーム・リレーのような従来のパケット交換技術と同様、クラス2の交換は、接続上ではなく、FCデータ・フレーム上で行われる。ノード間には専用接続は確立されず、各フレームは、使用可能なルート上でその宛先まで送られる。クラス2トラフィックにおいて輻輳（congestion）が発生した場合、その宛先に首尾良く到達するまで、フレームを再送信する。クラス3サービスは、1対多数の無接続フレーム交換サービスを定義し、配信保証や確認機構がないことを除いて、クラス2サービスと同様である。クラス3の送信は、確認を待たないので、クラス2の送信よりも高速である。しかしながら、送信がその宛先に到達しない場合、クラス3のサービスは再送信を行わない。このサービスは、承認を待つことができないがクラス1サービスを保証する程に時間に厳しくない、リアル・タイム・ブロードキャストに最も多く使用されている。また、フレームの損失が許されるアプリケーションにも使用されている。クラス4サービスは、接続を基本とするサービスであり、部分的帯域(fractional bandwidth)の保証およびレイテンシ・レベルの保証を提供する。

【0022】FC－3レイヤ、即ち、レイヤ325は、FC－PHレベルより高い上位レイヤ・プロトコルの通信サービスの共通集合を提供する，これらの追加サービスは、例えば、マルチキャスト(multicast)およびブロードキャストのデータ配信機構、1つ以上のターゲット・ノードが所与のイニシエータ・ノードに応答可能な「ハント(hunt)」群、ならびに多数の上位レイヤ・プロトコルおよびFC－PHの多重化を含むことができる。

【0023】FCPスタック300の最上位レイヤであるレイヤ330は、FC－4レイヤである。これは、例えば、図2に示したFC環境220のようなFCインフラストラクチャ上で動作可能な上位レイヤの用途を定義する。FC－4レイヤは、既存のチャネルおよびネットワーク・プロトコルを、ファイバ・チャネル上で、これらのプロトコルを変更することなく利用する方法を提供する。したがって、FC－4レイヤは、プロトコル収束レイヤ(protocol convergence layer)のように作用するので、FCノードは、上位レイヤのチャネルまたはネットワーク・プロトコルが要求する正確な下位レイヤ搬送サービスを提供するように見える，この収束機能は、FC－4レイヤが、バッファ記憶、同期、またはデータの優先順序付けのような、追加のサービスを提供することを、要求する。FC－4の機能は、図1に示した例示のコンピュータ・システム200の、FC環境220およびOS互換インターフェース215間に配されたリンク経路225に含まれる。

【0024】引き続き図2を参照すると、種々のFC－4レベル・マッピングが、多数の上位レイヤ・チャネルおよびネットワーク通信プロトコルのために指定されている。その中には、インテリジェント周辺インターフェース（「IPI」）、SCSI、高性能並列インターフェース（「HIPPI：High-Performance ParallelInterface）、単一バイト・コマンド・コード・セット（「SBCCS：SingleByte Commad Code Set）、論理リンク制御（「LLC：Logical Link Control」）、IP、および非同期転送モード（「ATM：Asynchrono

us Transfer Mode」）アダプテーション・レイヤ（「ＡＬ：Adaptation Layer」）が含まれる。

【００２５】ファイバ・チャネル・プロトコルと共に動作可能なデバイスは、それらがイニシエータであるか、あるいはターゲットであるかには無関係に、ＦＣＰスタック３００の中間レイヤのいくつかの機能を具体化するコントローラ（以降「ＦＣコントローラ」と呼ぶ）を通常含んでいる。例えば、現在のＦＣコントローラは、一般に、レイヤ３１５，３２０（ＦＣ－１，ＦＣ－２）の機能を具体化する。一方、図１に示した例示コンピュータ・システム２００のようなホスト・コンピュータは、上位レイヤ（ＦＣ－３，ＦＣ－４）を担当する。例えば、ギガバイト・リンク・モジュール（「ＧＬＭ：Gigabit Link Module」）のような物理的リンク・モジュール（「ＰＬＭ：Physical Link Module」）が、最下位レイヤ３１０（ＦＣ－０）を実現する。

【００２６】次に図３のＡ～Ｃを参照すると、３種類のトポロジ構成例が、全体的に４９０，４９１，４９２でそれぞれ示されており、その中にＦＣノードを配置することができる。ノードとは、ＵＬＰ、ＦＣ－３、およびＦＣ－２機能のいくつかを処理する能力を有する実体、システム、またはデバイスのことである。ノードは、１つ以上のポートを含むことができ、これらは一般的にノード・ポートすなわちＮ＿ＰＯＲＴとして知られている。Ｎ＿ＰＯＲＴは、ＦＣ－ＰＨを支援するノード内のハードウエア実体である。これは、送信元（即ち、イニシエータ）、応答側（即ち、ターゲット）、または双方として動作することができる。以降、ノード、デバイス、およびポートという用語は、本発明の目的のために、相互交換可能に用いることにする。

【００２７】図３のＡの４９０は２点間トポロジを示し、双方向通信リンク４１０Ａ，４１０Ｂを利用して、任意の２つのＦＣノード（ここではＮ＿ＰＯＲＴ４００Ａ，４００Ｂで示す）の間に全二重送信経路を提供する。この接続トポロジは、介在するデバイス／ノードがないので、可能な最大帯域および最低のレイテンシを提供する。

【００２８】図３のＣの参照番号４９２は切替ファイバ・トポロジであり、各ＦＣデバイス即ちノード（Ｎ＿ＰＯＲＴ）が、ファブリック、例えば、ファブリック４３０の一部であるＦ＿ＰＯＲＴに接続され、このファブリック上の他のいずれかの接続への無遮断データ経路を受ける。Ｆ＿ＰＯＲＴとは、物理的に他のノードに接続するためのファブリックのアクセス点である。ファブリック４３０は、スイッチまたは一連のスイッチとすればよく、ノード間のルーティング、エラー検出および訂正、ならびにフロー制御を担当する。ファブリック４３０の動作は、上位レイヤの通信プロトコルとは独立しており、専ら距離には無関係であり、いずれの技術を基準にしてもよい。

【００２９】通信経路、例えば、経路４３９は、ノード、Ｎ＿ＰＯＲＴ４４０、およびファブリックのポート（Ｆ＿ＰＯＲＴ）４３６間に双方向接続を提供する。切替ファブリック・トポロジ４９２は、３種類のＦＣトポロジ全ての内、最大の接続能力および総計スループットを与える。切替ファブリック・トポロジ４９２は、多数のシステムを相互接続し、高帯域に対する要求を支持し、異なる速度の接続間でデータ速度を一致させ、異なるケーブル要素の整合を取る機能を提供する。

【００３０】図３のＢの４９１は、ＦＣ－ＡＬ規格と呼ばれる接続規格に準拠し、仲裁ループ（「ＡＬ」）として当技術では知られている、ループ・トポロジを示している。ループ・トポロジ４９１は、例えば、Ｌ＿ＰＯＲＴ４２０Ａ～４２０Ｄのような複数のＦＣデバイス即ちノード（ループ・ポート即ちＬ＿ＰＯＲＴとして示す）を、単一方向リンク、例えば、リンク４２５Ａ～４２５Ｄを通じて相互接続する。したがって、この接続構成は、各デバイスがループ・トポロジ４９１を、送出側および受信側間において二点間接続として使用することを可能にするものである。その際、それらの間にどのようなデバイスが介在しても無関係であり、それらは単に「リピータ（中継器）」として作用するに過ぎない。

【００３１】仲裁ループ４９１は、ハブまたはスイッチを必要とせずに、多数のデバイスを取り付ける低コストの手段を提供する。図３のＢには４つのＬ＿ＰＯＲＴのみを示すが、このループは１２７個までのＬ＿ＰＯＲＴの共有帯域を提供する。各Ｌ＿ＰＯＲＴは、他のポートと通信する必要がある場合に、ループの使用を要求する。ループが未使用状態の場合、要求元のポートは宛先ポートとの双方向接続を設定する。ループ・プロトコルによって、Ｌ＿ＰＯＲＴは、連続的に送信媒体に対するアクセスのアービトレーション（仲裁）を行い、他のＬ＿ＰＯＲＴに送信することが可能となる。フェアネス・アルゴリズム(fairness algorithm)によって、ループへのアクセスが阻止されるＬ＿ＰＯＲＴが生じないことを保証する。一旦接続が確立されると、次に、２つのＬ＿ＰＯＲＴ間のトラフィックに適したいずれかのサービス・クラスで、配信が可能となる。

【００３２】当技術では公知のように、一度に通信可能なＬ＿ＰＯＲＴは１対のみである。これらのＬ＿ＰＯＲＴがループの制御を放棄した場合、他の２つのＬ＿ＰＯＲＴ間の二点間接続を確立することができる。更に、ループ全体は、ＦＬ＿ＰＯＲＴとして知られているものを通じてＦＣ交換ファブリック・ポートに、またはＮＬ＿ＰＯＲＴを通じて直接単一のホスト・システムに、交互に取り付けることも可能である。本発明の現時点における好適な例示実施例は、ループ・トポロジ４９１のようなＦＣ－ＡＬトポロジを含み、このノード構成の全体的な動作について、以下で更に詳細に説明する。

【００３３】ＦＣ－ＡＬ規格は、各ＦＣデバイスが、ル

ープ初期化プロセスにおいて、仲裁ループ物理アドレス（ＡＬ＿ＰＡ：Arbitrated Loop Physical Address）について取り決めることを許しており、これは知られていることである。仲裁ループに参加している間、ＦＣデバイスは、ループ・トランザクションを開始する前に、互いにログインしなければならない。ログイン手順は、全ての通信ノードが実行し、サービス・パラメータや共通動作環境を確立するための初期手順である。サービス・パラメータの例の１つに、「クレジット」限度がある。これは、受信側ポートにおいてバッファ記憶の溢れ(overrun)を発生させることなく、ポートが送信可能な未決フレームの最大数を表す。クレジットとは、各送信元ポート送出可能なフレーム数を限定することによって、リンクのトラフィックを絞るフロー制御機構であることは自明であろう。従来のＦＣコントローラでは、バッファ対バッファ・クレジット（「ＢＢ＿Ｃｒｅｄｉｔ：buffer-to-buffer credit」）および端末（エンド）対端末クレジット（「ＥＥ＿Ｃｒｅｄｉｔ：end-to-end credit」）という２種類のクレジットが一般に用いられている。

【００３４】あるデバイスが他のデバイスにログインしない場合、ログインするまで、当該デバイスから受信したあらゆるフレームをディスカード（破棄）する。イニシエータまたはドライバがそれが通信しているターゲット・デバイスを管理することができなければならないので、イニシエータは、当該ターゲット・デバイスに対するＦＣ－特定識別トリプレット(identity triplet)を追跡する。このＦＣ特定ＩＤトリプレットは、ターゲットのノード名、そのポート名、およびそのＡＬ＿ＰＡから成る。ＡＬ＿ＰＡはループ・リセット時に動的に割り当てられるが、ノード名およびポート名は、装置の一意のワールド・ワイド名(World_Wide_Name)から形成される。

【００３５】リセット時に、デバイスが仲裁ループ上に上った場合、デバイスは、ループ初期化ステップにおいて３つの方法の１つでデバイスのＡＬ＿ＰＡを環境設定する。ソフト・アドレス方式(Soft Address scheme)では、デバイスは、どのＡＬ＿ＰＡがそれに割り当てられるかには注意しない。むしろ、最初の未使用ＡＬ＿ＰＡで入手可能なものを単純に受け入れる。好適アドレス方式(Preferred Addressscheme)では、ＦＣデバイスは、特定のＡＬ＿ＰＡが割り当てられることを望む。しかしながら、何らかの理由で所望のＡＬ＿ＰＡが利用できない場合、未使用でありかつ入手可能ないずれかのＡＬ＿ＰＡを受け入れる。例えば、ＯＳのロードに続く「全体的な」システムの初期化時に、デバイスに特定のＡＬ＿ＰＡが最初に割り当てられた後、このデバイスは、後続のループ・リセット時に、当該ＡＬ＿ＰＡを要求し続ける。しかしながら、一旦このデバイスが仲裁ループから外れたなら、そのＡＬ＿ＡＰを「選択」する能力は失われ、入手可能な最初の未使用ＡＬ＿ＰＡを受け入れことに甘んじなければならない。

【００３６】第３に、ハード・アドレス方式(Hard Address scheme)では、ＦＣデバイスは、特定のＡＬ＿ＰＡでのみ動作することができる。ＡＬ＿ＡＰの環境設定を扱うＦＣ－ＡＬ規格におけるループ初期化プロトコル（「ＬＩＰ」）によれば、このアドレス構成方法は、最初の２つの方法、即ち、ソフト・アドレス方式および好適アドレス方式に勝るものである。

【００３７】イニシエータＦＣデバイスは、全てのＡＬ＿ＰＡ割り当て問題が解決した後に、リンク・サービス・コマンド／フレームを開始することができる。リンク・サービス・フレームは、「要求」フレームおよび「応答」フレームの双方を含む。要求フレームは、受信側デバイスに、応答フレームを返送するように要求するリンク・サービス・フレームであり、例えば、ログイン・リンク・サービス・フレーム（「ＰＬＯＧＩ：Login Link Service Frames」）、ログアウト・フレーム（「ＰＬＯＧＯ：Logout Frames」）、Ｎ＿ポートサービス・パラメータ発見フレーム（「ＰＤＩＳＣ：Discover N_Port Service Parameters」）、アドレス発見フレーム（「ＡＤＩＳＣ：Discover Address Frames」）、プロセス・ログイン・フレーム（「ＰＲＬＩ：Process Login Frames」）、プロセス・ログアウト・フレーム（「ＰＲＬＯ：Process Logout Frames」）、および復位復元資格フレーム（「ＲＲＱ：Reinstate Recovery Qualifier Frame」）を含んでいる。

【００３８】単一イニシエータ環境では、イニシエータ・デバイスは、必要に応じて、リンク・サービス・フレームを送出し、それに応答して、承認フレーム（ＬＳ＿ＡＣＣ）または拒絶フレーム（ＬＳ＿ＲＪＴ）を期待する。更に、イニシエータ・デバイスは、未決リンク・サービス・アレイと呼ばれる記憶アレイ内に、各リンク・サービス・フレームに対するタイプ情報（以降、「タイプ情報要素」と呼ぶ）を記憶することにより、送出されるリンク・サービス・フレームのタイプを追跡する。この未決リンク・サービス・アレイは、複数の記憶位置から成り、その各々が受信側デバイスのＡＬ＿ＰＡに対応する。更に、典型的な実施例では、全てのリンク・サービス・フレーム・タイプは、それらが送出される際に、各受信側に記憶される。

【００３９】イニシエータ・デバイスによる初期ポート発見プロセスは、それが１つ以上のイニシエータを含んでいるか否かには無関係に、ＦＣ－ＡＬ環境では二段階プロセスである。最初に、イニシエータが既に受信側デバイスにログインしている場合、ＰＤＩＳＣフレームが送信される。その他の場合、ＰＬＯＧＩフレームが送信される。第２に、ＰＬＯＧＩフレームに応答してＬＳ＿ＡＣＣフレームが返送され受信された場合、イニシエータは、当該応答側にＰＲＬＩフレームを送る。一方、Ｐ

ＤＩＳＣフレームに応答して、ＬＳ＿ＡＣＣフレームが返送され受信された場合、他のフレームを当該応答側に送る必要はない。

【００４０】マルチ・イニシエータ環境では、イニシエータがそれ自体のＰＬＯＧＩ要求フレームに応答してＬＳ＿ＡＣＣまたはＬＳ＿ＲＪＴフレームを期待している場合、ＰＬＯＧＩフレームがイニシエータに送られる。２カ所のイニシエータが互いに通信する必要がある場合、これらは双方ともログイン・プロトコルを継続しなければならない。何らかの理由で、一方のイニシエータが他方との通信を支援しない場合、送出側から受信するＰＬＯＧＩフレームを単に無視すればよい。即ち、ＬＳ＿ＲＪＴフレームを送出側に送ればよい。しかしながら、このようにすると、通信システムの性能を激しく低下させると共に、それに関連するスループットに負の影響を与えてしまうことは明らかである。

【００４１】次に図４を参照すると、本発明の教示による、２つのイニシエータ・デバイス、即ち、イニシエータ５１０およびイニシエータ５１５、ならびにターゲット５２０から成る仲裁ループ５００の例が示されている。イニシエータ５１０は、そのノード名称であるＮＯＤＥ＿Ａを用いてＡＬ＿ＰＡＩを処理する。イニシエータ５１５はそのノード名称であるＮＯＤＥ＿Ｂを有するＡＬ＿ＰＡ２に配置される。更に続けると、ターゲット５２０はＮＯＤＥ＿Ｃを有するＡＬ＿ＰＡ５に配置される。この図ではポート名を具体的に示していないが、それらは、デバイスの各々に関連するＩＤトリプレットに対して存在することは理解されよう。

【００４２】イニシエータ５１０は、未決リンク・サービス（「ＯＬＳ：outstanding_link_services」）アレイ５１２と呼ばれる記憶アレイを備えており、一方このアレイ５１２は、それぞれ、ＡＬ＿ＰＡ２（イニシエータ５１５に対して）およびＡＬ＿ＰＡ５（ターゲット５２０に対して）に送られる、あらゆるリンク・サービス・フレームにも対応するタイプ情報要素を格納する位置５１１Ａ，５１１Ｂを含む。同様に、イニシエータ５１５も、位置５１６Ａ（ＡＬ＿ＰＡ１、即ち、イニシエータ５１０のため）および５１６Ｂ（ＡＬ＿ＰＡ５、即ち、ターゲット５２０のため）を有するＯＬＳアレイ５１７を備えている。

【００４３】首尾良くループ初期化プロセスを完了した場合、イニシエータ５１０はＰＬＯＧＩフレームをイニシエータ５１５およびターゲット５２０の各々に送出する。イニシエータ５１０は、これらのＰＬＯＧＩフレームに対応するタイプ情報要素を、そのＯＬＳアレイ５１２内の位置５１１Ａ，５１１Ｂに記憶する。同様に、イニシエータ１５５は、ＰＬＯＧＩをイニシエータ５１０およびターゲット５２０に送出し、対応するタイプ情報要素をそのＯＬＳアレイ５１７内の位置５１６Ａ，５１６Ｂに記憶する。

【００４４】イニシエータ５１０がイニシエータ５１５からＰＬＯＧＩフレームを受信した場合、例えば、ＬＳ＿ＡＣＣフレームのような応答フレームをイニシエータ５１５に送出する。本発明の教示によれば、イニシエータ５１０は、位置５１１Ａに対応するタイプ情報要素を記憶することなく、単に応答フレームを送出する。したがって、位置５１１Ａは、イニシエータ５１０によってイニシエータ５１５に送出されたＰＬＯＧＩフレームに対応する、最初のタイプ情報要素を保持する。このため、応答フレーム（例えば、ＬＳ＿ＡＣＣ）がイニシエータ５１５からイニシエータ５１０に返送されたとき、イニシエータ５１０は、受信した応答フレームは、事実上、イニシエータ５１５に送出されたそれ自体のＰＬＯＧＩフレームに応答するものであることを、素早く判断することができる。

【００４５】例えば、イニシエータ５１５がイニシエータ５１０からのＰＬＯＧＩフレームに応答するよりも前に、イニシエータ５１５からのＰＬＯＧＩフレームがイニシエータ５１０に到達し、イニシエータ５１０がそのＬＳ＿ＡＣＣによって応答し、そのタイプ情報要素を位置５１１Ａに記憶した場合、イニシエータ５１５からのＬＳ＿ＡＣＣフレームを受信する際に、イニシエータ５１０はログイン・プロセスに素早く進むことができない。何故なら、イニシエータ５１５からＬＳ＿ＡＣＣフレームを受信した後、受信したＬＳ＿ＡＣＣのペイロードを分析し、「見失わない」ようにしなければならないからである。更に、ＬＳ＿ＡＣＣフレームに対するタイプ情報要素がＯＬＳアレイ５１１に記憶されるときに返送されたＬＳ＿ＡＣＣを受信する際、イニシエータ５１０は中間状態に入ることによって、ログイン・プロセスが止まってしまう場合がある。しかしながら、本発明の教示にしたがって多数のイニシエータを設けることにより、かかるループの中断を大幅に抑え最少化する。

【００４６】図５は、例えば、先に述べた仲裁ループ５００のようなファイバ・チャネル環境における多数のイニシエータを制御する、現時点における好適な方法のフロー・チャートを示している。ステップ６０１においてループ初期化を完了したときに、まず、イニシエータは、発見された有効なＡＬ＿ＰＡ全てに送信される、適切な要求フレームに対するタイプ情報要素を記憶する。このプロセスはステップ６０５に示されている。続いて、ステップ６１０に示すように、適切な最初の即ち第１の要求フレームを送信する。他のイニシエータ・デバイスから要求フレーム（第２の要求フレーム）を受信した場合、そのタイプ情報要素を記憶することなく、適切な応答フレームを返送する（ステップ６１５）。引き続き図５を参照すると、適切な応答フレームをイニシエータから受信した場合、受信したイニシエータは、その内容を操作することなく、この応答フレームを適切に処理する（ステップ６２０）。かかる処理は、受信した情報

の読み取り、質問等を伴う場合があるが、受信した応答フレームに関連するペイロードの分析を必要としないことが好ましい。

【００４７】以上の説明から、本発明により、ＦＣ環境において多数のイニシエータの制御および管理を行う画期的な機構を提供することによって、従来技術の問題を首尾良く解決することができることは、当業者に明らかであろう。本発明のいくつかの実施例のみを添付図面に示し前述の詳細な説明に記載したが、本発明は開示した実施例に限定される訳ではなく、特許請求の範囲に記載しそれによって規定される本発明の精神から逸脱することなく、多数の再構成、変更および交換が可能である。

【図面の簡単な説明】

【図１】本発明を実施可能な、コンピュータ・システムの一例を示すブロック図である。

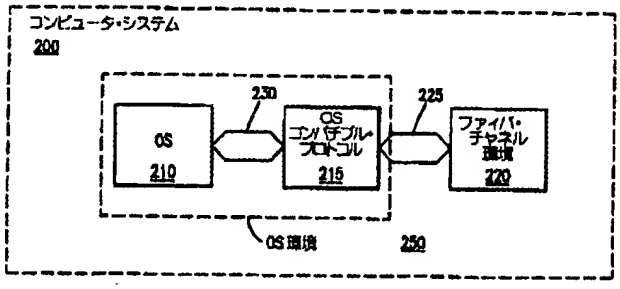
【図２】ファイバ・チャネル（ＦＣ）プロトコル・スタックを示す模式図である。

【図３】ファイバ・チャネル・ノードに使用可能な３種類のトポロジ構成を示すブロック図である。

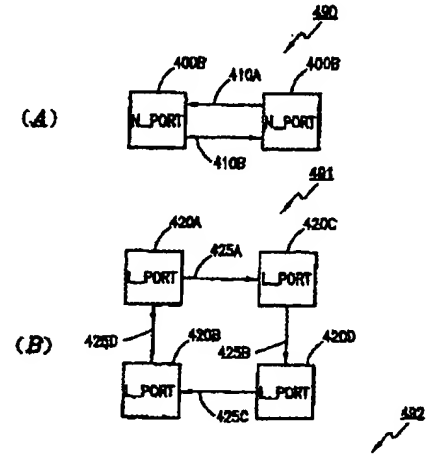
【図４】本発明による、多数のイニシエータを有する仲裁ループの代表的実施例を示す図である。

【図５】本発明にしたがって、マルチ・イニシエータの仲裁ループを制御する方法を示すフロー・チャートである。

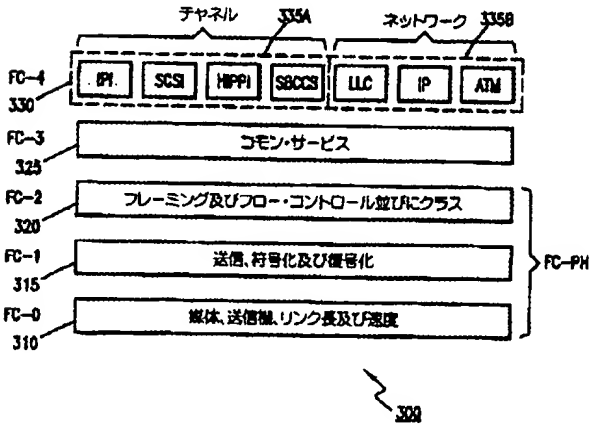
【図１】



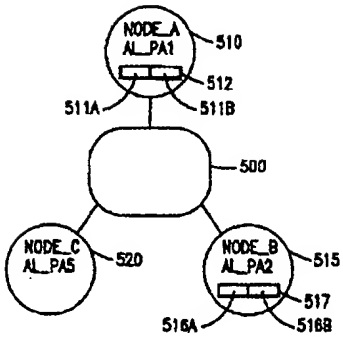
【図３】



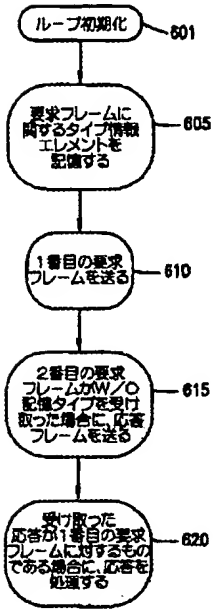
【図２】



【図4】



【図5】



フロントページの続き

(71)出願人　591030868
20555 State Highway
249, Houston, Texas
77070, United States o
f America